

## **Minutes of Federal Gazetteer Users Meeting**

USGS Headquarters, Reston, VA

22-23 May 2007

### **Meeting Notes of the Federal Gazetteer Users Meeting**

**May 22-23, 2007**

### **US Geological Survey Headquarters, Reston, Virginia**

Notes compiled by Denise Bleakly from notes taken by Marvin Lindroth, NGA, Judy More, Sandia National Labs, and Denise Bleakly, Sandia National Labs.

The **Federal Gazetteer Users Meeting** was held at the US Geological Survey Headquarters in Reston, Virginia, on May 22-23, 2007. It was hosted by the US Geological Survey (USGS), the National Geospatial Intelligence Agency (NGA), and the PATTON Alliance (Sandia National Laboratories).

Representatives from various users groups included: DIA, CIA, NSA, Dept. of State, DHS, NGIC, NASIC, FMSO, and Library of Congress met with representative of the NGS and USGS to discuss the future needs for gazetteers, place name services and "place name intelligence".

#### **Day 1, Tuesday, 22 May 2007**

**Denise Bleakly**, Gazetteer Project Lead of Sandia National Laboratories, opened the Meeting. She is with Sandia National Laboratories which is home to the PATTON Alliance, whose acronym means "Preparing Analytic Tools and Technology for an Operational Network." PATTON is a technology middleman to get "government-funded technology ready for integration and deployment to specific end users," which in this case means the US government.

**Denise Bleakly** stated in the written agenda that the purpose of the conference was *"to bring together PATTON Alliance Members, with NGA, USGS and other federal stakeholders to discuss issues concerning the use of gazetteers in support of the Intelligence Community (IC), to identify needs and explore how to close the gaps between user needs and current gazetteers."*

The goals of the conference were the following:

1. Understanding the role that NGA and USGS play in the identification of geographic place names.
2. Identifying the gaps between a government-funded research gazetteer and what is needed in a fielded government system.
3. Determining if a PATTON-identified technology will meet the needs of fielded federal government systems.

**Jennifer Schnarre**, chief of NGA/PR, welcomed the conference participants on behalf of NGA. She presented her goals of the conference:

1. the opportunity to exchange information
2. the discussion of future trends for gazetteers
3. a forum for the needs of gazetteer users

**Jennifer Schnarre** hoped that information gaps in gazetteers based on users' evolving needs will be filled and that such conferences will be held annually.

**Lou Yost**, USGS and Executive Secretary of the Domestic Names Committee (DNC) of the US Board on Geographic Names (BGN), welcomed everyone to the US Geological Survey (USGS). He mentioned that the conference will be tilted toward foreign place-name gazetteers. His agency, the USGS, produces the Geographic Names Information System (GNIS), which is the official digital gazetteer of US place names.

## Minutes of Federal Gazetteer Users Meeting

USGS Headquarters, Reston, VA

22-23 May 2007

**Judy Moore**, Program Manager the PATTON Alliance at Sandia National Laboratories, welcomed conference attendants on behalf of the PATTON Alliance. She stated that members of the PATTON Alliance are technology developers reviewing the transition of new technology from research to deployment. The Gazetteer Project is aimed at searching for existing technology and determining current needs for gazetteers. Current funding for their gazetteer evaluation project ends in July 2007.

**Randy Flynn**, NGA and Executive Secretary of the Foreign Names Committee (FNC) of the BGN, represented the BGN.

Randy Flynn gave an interesting context setting presentation. In it he pointed out that they are really in the business of **place name intelligence**. Their users expectations are high:

- Adherence to standards
- Ease of access
- Toponymic information which meshes seamlessly with other information
- The right content

Their past successes have occurred in mathematical space. Their new challenge is precision identification in human space.

The challenges facing them have to do with:

- Scale
- Diversity of content
- Colloquial toponymy
- Ambiguity
- Deeper analytic capabilities

Their services need to include:

- Non-English, non-Roman alphabets
- Non-textual data
- Place name trolling
- Conflation/deconfliction
- Contextual confidence indicators
- Virtual collaborative collection

In addition, He discussed the traditional and historical use of gazetteers, and that users have expected the BGN to gather and compile names. He brought up the concept of **place-name intelligence**, which brings new challenges for future gazetteers.

**Trent Palmer**, NGA and Deputy Executive Secretary of the FNC, gave an overview of the role of the Board on Geographic Names (BGN). He encouraged users to offer their research and advice to BGN to improve and expand gazetteer data.

Here are some of the challenges that the FNC faces:

- ❖ resistance to change
- ❖ lack of sources
- ❖ staff shortage
- ❖ larger scope of gazetteer needs

**Lou Yost**, USGS and Executive Secretary of the Domestic Names Committee (DNC), discussed the role of the DNC. Members of the DNC include representatives from the following agencies:

## Minutes of Federal Gazetteer Users Meeting

USGS Headquarters, Reston, VA

22-23 May 2007

- ❖ Departments of Agriculture
- ❖ Commerce
- ❖ Interior
- ❖ Homeland Security
- ❖ Government Printing Office
- ❖ Postal Service
- ❖ Library of Congress.

The committee goes over approximately 400 name proposals per year and enters approximately 3000 names into the names database (GNIS) per month.

There are 50 state names authorities that form the Council of Geographic Names Authorities (COGNA).

**Dwight Hughes**, USGS, introduced the topic of *The National Map*, which covers the entire country with seamless USGS maps. The names come from the GNIS, which contains over 2 million features.

**Barbara Locke**, NGA, introduced the future Geographic Names Server (GNS) that stores the foreign-names database of the US Government. She mentioned that the future GNS is an Oracle database rather than a Sybase database. The future GNS is part of the GGMA initiative for an integrated data-centric environment, i.e. a conflation of digital databases.

**Fawaz Alami**, NGA, demonstrated the future GNS website.

**Peter Viechnicki**, Senior Linguist at NGA, discussed future trends in gazetteer production and maintenance. He wants to bring automated processes to ease the work which has been mostly entered manually and, therefore, labor intensive and time consuming. Among the automated processes are the following:

- ❖ Optical Character Recognition (OCR) for lists of names
- ❖ Transliteration
- ❖ GeoDDupe, which is a gazetteer conflation tool being worked on by the University of Maryland
- ❖ MITRE's Geospatial Data Integration Algorithms Tool for gazetteer file matching
- ❖ NGA's intellimapia – Peter wants to have a user-contribution website by 2008, in which users contribute names data which is similar to the existing Wikimapia, which consists of over 3.5 million user-contributed names.
- ❖

**Quinton Congdon**, National Ground Intelligence Center (NGIC), spoke about the use of gazetteer information for intelligence work. NGIC analyses foreign information of interest and needs the gazetteer information to plot place names from its information onto maps. He would like to see improved coordinates in gazetteers. A lot of NGIC's gazetteer information comes from the GNS.

**Deborah Tatarinchik**, NSA, indicated that NSA is a heavy user of GNS. Unfortunately, many features, such as buildings and streets, are not in the GNS, and, consequently, she is tasked to do the research of these missing features. She envisions an improved gazetteer as one that includes named features found on large-scale maps and has precise coordinates. She would also like to see the capability of proximity and buffer searches and other ArcGIS functions, i.e. an ArcGIS-based gazetteer.

NSA needs:

- Entities below Admin1
- Multiple place names associated with a given place
- Neighborhood details in large cities
- Street addresses in foreign countries

## Minutes of Federal Gazetteer Users Meeting

USGS Headquarters, Reston, VA

22-23 May 2007

- Variant names, multiple languages
- Place names misspelled, bad transliteration, translated by other nations
- Large, prominent building in foreign cities
- Transportation facilities to include ports and other modes of transport
- More precision in coordinates down to seconds in decimal degrees
- Proximity search and radius search in GNS
- Have available a gazetteer within ArcGIS
- They use Google Earth but mostly ArcGIS

During Deborah's presentation, *Google Earth* as a visualization tool was mentioned, and another visualization tool called *World Wind* by NASA was brought up by members of the audience.

**Allan Wiley**, Army Corps of Engineers, Topographic Engineering Center (TEC), informed us that his organization is also a big user of GNS. He would like to see more attributive data in the GNS, such as elevation and population.

**Leo Dillon**, Office of Geography, Department of State, and FNC member emphasized that gazetteer information is constantly needed. His office receives many inquiries from Foreign Service personnel abroad that can be answered by searching in the GNS. The people, however, find the GNS too awkward to use, and therefore prefer to call his office. Leo recommends strongly that the GNS must be easier to use, especially for the unsophisticated user. Codes should be replaced by clear, spelled-out information, such as country names and feature designations. GNIS, which is the US names gazetteer, is easier to use because it is intuitive. The GNS should also contain links to other sources of data so additional data, even photographs, can be found. The GNS search capability in other languages and scripts would also be useful. Also, the returns of a name search should be the most likely match should ranked with the first on the list of returns be the most used or most likely match.

**Bohdan Kantor** and **Michelle Rego**, Library of Congress (LoC), spoke about LoC's prototype project, the *World Digital Library*.

As other organizations, the project relies heavily on gazetteer data. Some of their issues about gazetteer data are the following:

- ❖ Poor data
- ❖ Inconsistency in metadata
- ❖ Getting data to adhere to standards of foreign libraries involved in the project
- ❖ Data file conflation

After the presentation by LoC, a discussion started concerning the definition of *neighborhood*. USGS does not use *neighborhood* in its gazetteer, because it is a vague term.

Also, a discussion was concerning separate uses of gazetteer data emerged. People end users, such as those at the Department of State or NSA, versus the "machine users" where a gazetteer data set was needed as part of a larger system for many uses. The "people end-users" needed better user interfaces that were more intuitive, more user friendly, while the "machine end users" needed better data structures, data models, standards, and data dictionaries.

## **Minutes of Federal Gazetteer Users Meeting**

USGS Headquarters, Reston, VA

22-23 May 2007

### **Day 2. Wednesday, 23 May 2007**

Today's discussions are about the future direction of gazetteers. The participants are unanimous about the need for gazetteers to improve and fill gaps that users need.

The day was designed to have two speakers, Frank Linton, of MITRE Corporation who did a study of end users' experience of GNS, and Denise Bleakly, and her experience in collecting user requirements for gazetteers for intelligence systems.

After these two presentations, a discussion was facilitated by Judy Moore of Sandia National Laboratories, to list action items and next steps.

**Frank Linton**, MITRE, wrote a report in 2005 about the GNS and suggestions for improving it. The suggestions came from interviews with many GNS users.

Frank mentioned that the GNS is "painful" to use. He showed the GNS website and explained why it was so. He said that it was difficult to understand the content without a lot of reading and studying. Here are other comments:

- ❖ codes were not intuitive
- ❖ population data is missing
- ❖ scale selection is not available
- ❖ links to additional sites for information are not available

Frank recommended some changes to the GNS:

- ❖ users should be encouraged to contribute names information
- ❖ news of data updates should be available
- ❖ dates of names data and their sources should be given
- ❖ allow users to select the data that they need
- ❖ offer various formats for downloading
- ❖ provide phonetic aids
- ❖ improve wildcard searches
- ❖ prevent searches to time out
- ❖ provide alternative coordinate systems
- ❖ search in foreign script
- ❖ make GNS easier to find online

As expected, Frank's presentation evoked a passionate discussion. Here are some points:

- ❖ As a start, the user interface should be corrected first before going on to the gazetteer's content and form.
- ❖ There needs to be a clear definition of gazetteer. What is its purpose? What should it contain? The gazetteer must evolve to fulfill current and future user needs.
- ❖ The GNS is problematic for the user perhaps because it contains foreign data that users are unfamiliar with.
- ❖ Provide users a template of a data model so that they can produce their own specific gazetteer.

**Denise Bleakly**, Gazetteer Evaluation Project Lead, Sandia National Laboratories, and presented the PATTON Alliance Gazetteer Evaluation Project. The PATTON Alliance members are looking at the system level to make improvements to the current gazetteers. She mentioned that some foreign governments have their gazetteers online and are good sources of data. The problem, however, is how the data can be integrated with our gazetteers. While striving for the ideal gazetteer, problems must be

## Minutes of Federal Gazetteer Users Meeting

USGS Headquarters, Reston, VA

22-23 May 2007

solved in turn rather than trying to solve all problems before releasing a new gazetteer. The expectations of the gazetteer have evolved quickly in the high-tech reality, surpassing its original intent of providing a list of names, description, and geographic location.

**Judy Moore**, Project Manager, Sandia National Laboratories, guided the group discussion further by offering five categories in which we can write ideas for improving gazetteers:

- ❖ Content
- ❖ Standards
- ❖ Integration
- ❖ Interface & Access
- ❖ Other

Content ideas:

- ❖ Search ability
- ❖ Unique Name ID
- ❖ Name with & without diacritics
- ❖ Single-feature classification in GNS and GNIS
- ❖ Point and bounding box for coordinates
- ❖ Position in hierarchy
- ❖ Multilingual
- ❖ Multi-lingual script
- ❖ Phonetics
- ❖ Official abbreviations
- ❖ Adjective forms
- ❖ Temporality (time frame for names)
- ❖ Various file formats
- ❖ Map files
- ❖ Image Files
- ❖ Data sources for measure of reliability
- ❖ Expanded Feature Classes (e.g. critical infrastructures)
- ❖ Population Data
- ❖ Elevation
- ❖ Reliability/accuracy measures

Standards ideas:

- ❖ Unicode (use it)
- ❖ Transliteration (use these standards more)
- ❖ Standard gazetteer data model needed
- ❖ Data-exchange formatting standard
- ❖ Glossary of terms

Integration ideas:

- ❖ Merge GNIS and GNS
- ❖ Foreign official datasets
- ❖ Normalization
- ❖ Importing and exporting data
- ❖ Conflation tools
- ❖ Abstractions to allow integration

## **Minutes of Federal Gazetteer Users Meeting**

USGS Headquarters, Reston, VA

22-23 May 2007

- ❖ Software services to assist in integration
- ❖ How to integrate from “shoe boxes” (distributed model)

### **Interface & Access ideas:**

- ❖ Web query
- ❖ Multiple-format downloading
- ❖ Web services
- ❖ Mirroring
- ❖ Public & Classified access
- ❖ Findability (how easy/hard to find the gazetteer data/service)
- ❖ Usability (define classes of users)
- ❖ Interface to the Help file

### **Other ideas:**

- ❖ Need for use cases
- ❖ (Who will get these requirements?)
- ❖ Data collection from non-traditional sources – reliability?
- ❖ Prepare for a BILLION place name entries in the near future
- ❖ Archiving
- ❖ Metadata
- ❖ Data conversion from other sources
- ❖ Web harvesting

### **Work-approach priorities:**

- ❖ Build working groups for each of the five categories
  - ❖ Scan archived and microfilmed names cards for their wealth of additional information
    - Scan NGA index cards to get information out to the larger world – analysts
    - MITRE has some technology
    - Carl has some ideas
    - BGN index cards?
    - BGN case files?
    - Library of Congress place name authorities.loc.gov
  - ❖ The standards group needs to set up standards that need to change
  - ❖ The other groups need to set up requirements. The requirements need to be identified to justify the resources needed for the gazetteer project
  - ❖ The PATTON Alliance has a SharePoint site. It can be accessed by anyone with a user ID and password. These are set up by the PATTON Alliance.
  - ❖ Working groups should hold their first meeting by July 2007. NGA will chair each group.
  - ❖ The next Gazetteer Users conference should be a full two-day conference held next year in May.
1. It was decided to formulate several working groups to develop requirements and make recommendations to the government for meeting the needs identified in this conference. The plan is to convene the first set of working group meetings before the end of July, 2007.
  2. Sandia will work with NGA to make sure these first meetings are held and the notes from this meeting are provided to these groups.

## **Minutes of Federal Gazetteer Users Meeting**

USGS Headquarters, Reston, VA

22-23 May 2007

3. Sandia will also grant access to the PATTON SharePoint site as a starting point for there working groups. NGA has agreed to provide a leader for each working group.
4. It was also decided to make this conference an annual event. NGA and USGS will take on the responsibility to host this conference as an annual event. They will consider having it in this same timeframe or placing it adjacent to another larger Geo conference like GEOINT. An opportunity for vendor displays or demos will be included. We noted that we need to invite DTRA to this meeting also.
5. The presentations from this meeting will be on the sharepoint site next week (5/30/07 and anyone here wanting access should contact Judy Moore (505-845-9415). The raw notes will be sent to all participants next week.